

## Summary Notes on the Talk with Demos

**Melitte Buchman [presenter] & Rick Ochoa**

**“Video Preservation Workflow at NYU – Performance on the Edge”**

We’re presenting where we are now with video preservation at NYU, what we’re actually producing, and how we’ve thought about it.

[“Overview” slide] This is an overview of R&D starting 18 months ago, when the landscape significantly differed from today. We’ll outline the adopted preservation workflow and goals, and define our understanding of preservation, which differs from that of industry. An important unanticipated part of the project is the importance of creating resident expertise, particularly in video, which was new to many of us.

[“Hemispheric Institute” Slide; bring up web site home page, [hemi.nyu.edu](http://hemi.nyu.edu)] The project involves work with the Hemispheric Institute, our graduate school of performance and politics. They had created a nice access web site as a teaching tool, but found they had a core of unique, ephemeral material that needed preserving. Only materials of “permanence” can go to the NYU Library.

[“System Overview” Slide – photo of racks, computer, and student] Our system started out as a Mini-DV deck, Digital BetaCam deck, a Mac G4, and a Medéa RAID array. We struggled with this for 12 months.

We then partnered with our Preservation Department, whose head, Paula De Stefano, was broadminded enough to run analog and digital things through the same system. By combining resources, we now have good calibration tools

Here’s Lindsey making master tapes, in front of the racks.

[“Hardware Overview” Slide

Three racks

Main System

- Sony DVW A500 DigiBeta VCR
- Sony OSR-1500GA DV VCR
- Power Mac G5 (4GB RAM, 2 x 250 SATA Disks RAID 0; 2x1TB RW800 LaCie“Big Disk”, Digital Voodoo RT64 PCI Card)

Calibration & Adjustment

- Tektronix TG700 Signal Generator
- Tektronix 611A Waveform Monitor
- Snell & Wilcox Time Base Corrector

Miscellaneous

- Serial Digital Interface Video Switcher
- Horita Longitudinal Time Code Generator
- Other Stuff

]

The heart of the system is the DigiBeta deck. (The Preservation Department also uses the BetaCam SP deck, but not us). There's a series of tools - signal generator, waveform monitor, time-base corrector.

[“Software Overview – Basically All Apple, All the Time” Slide] While hardware is king, software is also important. A lot of testing was done with non-integrated suites and separate pieces of software. But Apple's production suite works best for us, because it's well integrated:

- Mac OSX 10.3.X
- QuickTime 6.5.x
- Final Cut Pro 4.5.x HD
- Compressor 1.2.1
- DVD Studio Pro x

There have been problems with hardware and software going extinct while we were testing them. This remains the downside of video: the landscape changes quickly, forcing awareness of business models.

[“Digital Format Comparison Table” slide – approximated in table shown]

Format	Bit Rate, Mbps	Signal Sampling	Compression Ratio	Tape width, inches
DVCPRO	25	4:1:1	5:1	¼
DVCPRO 50	50	4:2:2	3.3:1	¼
DV & Mini DV	25	4:1:1	5:1	¼
DVCAM	25	4:1:1	5:1	¼
DIGITAL-S	50	4:2:2	3.3:1	½
Betacam SX	18	MPEG-2 Inter-Frame	10:1	½
Digital Betacam	90	4:2:2	2:1	½

18 months ago, HD and 4:4:4 hadn't raised their heads much. We felt, after evaluating our collection, with digital, what's listed on the slide would be the fanciest we would see in the near future. With DigiBeta tape, 4:2:2, for many inputs it's “pouring a quart of signal into a gallon container”, so a safe way of handling the signal.

[“Format Selection” slide] Tape was chosen for storage because of cost (\$50-70/94 minutes). Today's cost of our failover server for tape backup is \$100/hour, which is getting close to the tape rate. We expect crossover around 2010, a bit after Glenn's projection [from intro], but similar, at which time we expect to move material from tape to hard disk.

Our preservation standard is defined as “saving the SD signal”. The industry talks about saving the MPEG2 I-frames. We're more “library-centric”, concerned about primary sources. We look at the SD signal like a handwritten letter from George Washington. Future researchers may want to do something with the signal that we can't anticipate.

[“4:2:2 Workflow” Slide] The workflow is complete, running end-to-end, and producing assets. Our great new rack temporarily interrupted production and gave us new problems to solve.

[“4:2:2 Workflow – Calibration” slide] Our immediate goal is saving 250 hours/year, which should be no problem, probably exceed this.

This is our calibration workflow. Source material goes into an SDI switcher, and waveform monitor.

[Slide of waveform monitor screen] There’s a little bit of luma and chroma error, but basically we’re “hitting our points”. We weren’t at the beginning of the process, so the engineers were called in to rec.

[Slide] The signal is taken through the time-based corrector, through the DA, then to the waveform monitor.

[Slide of waveform monitor screen] Once were pleased, signal is sent on to preservation master. It’s RGB labeled, hitting all the points, and not getting any error. This is a great enhancement of the previous system of “hoping the calibration was correct”.

Bars and tone [shown] are put in front of every master tape, and the signal is pulled back and forth to make sure none is lost.

[“4:2:2: Workflow – Preservation Master” Slide] The DigiBeta deck has a “confidence head”, like a playback head in old machines. This allows viewing the source and master at the same time with a single monitor. It cuts 1-fold realtime off of the quality-assurance time, compared with separate viewing of source and master for damage.

[“Production Timeline” slide] Originally, 1 hour of video took 10 hours of production, little of which was passive. The confidence head reduced this to 1 to 9, and Rick is now trying to parallelize the workflow, by bringing derivative processing to a separate machine, and taking advantage of some passive time. [another slide briefly with itemized time.]

[“4:2:2 Workflow - Derivative Generation” slide] The workflow generates MPEG2 for DVD. Two DVD copies are made - one for our multimedia library, one for the artist or Hemispheric Institute. Two bandwidths (high and low) of MPEG4 are also created, but low, crafted for modem delivery, is being phased out.

[“HD Workflow” slide] In the future, the HD workflow is expected to be similar, but with greater processing needs. The biggest issue: how to store enormous files? Even with cheap storage,  $\frac{3}{4}$  of TB per hour is a lot to demand. What kind of tape could one take this down to?

[Access Beta sites, links to HI web site] Project member Eric Steadfield has worked on metadata needs and info presentation. What he's working on now for release shortly: Go to the "Bobcat" catalog record, find an asset of interest, click its URL, and get this panel links to various bandwidths [plays performance clip with Astrid Hadad. See also: [www.astridhadad.com](http://www.astridhadad.com)]. Students will be able to do this from dorm-room, as many students prefer. Will we need to do Palm next?

[METS Multimedia demo - Example training video "How to scan video".] Many assets are multimedia. Eric used SMIL to encode text. This will be useful if there's funding to encode text involving translation.

[Another video] As a final example, Hemispheric asked for something to be used in teaching, an interpretative use rather than archival. [There's an artist, Nao Bustamante, talking in an interview about a performance piece, with (if we had waited long enough) the piece shown beneath.]

["Metadata" slide] We've done a lot of work with metadata. We're working on a DSpace conversion, so the vehicle being used for technical metadata has to be transformed.

["Hemispheric Institute" slide, web demo] Currently in daily use is a web-based "Video Project Survey Form" between a subject specialist at the Hemispheric Institute and the catalogers, from which the MARC record derives. It uses a controlled vocabulary whenever possible, or at least hints to the Hemispheric Institute about what would be helpful to the cataloger. It maps somewhat to MARC, which may or may not be useful for every future purpose.

[clip, "Cosmic Blood", by Gigi Otálvaro-Hormillosa. See similar clips on artist's site: [www.devilbunny.org/mestizaje.htm#](http://www.devilbunny.org/mestizaje.htm#)] Showing the SD signal itself, this gigantic file (720 x 486 4:2:2 uncompressed) may choke the system [played locally from the Mac laptop], but is of beautiful quality. Gigi's artist statement discusses hybrid sexuality and her experience being a Filipino.

These are unusual and ephemeral works. A lot of artists are from South American countries, and sometimes the tapes received are too moldy to convert. It's important to not further postpone conversion.

[clip, "America the Beautiful", by Noa Bustamante - large, mostly-nude woman climbing a stepladder. See also a photo at artist's site: [www.naobustamante.com/works.html](http://www.naobustamante.com/works.html)] Her work often deals with issues of female beauty. A fearless performer, at the end of the piece she lights what remains of her costume (mainly a ribbon) on fire, at the top of the ladder.

[clip from the latin musical, "El Bolero fue mi Ruina" (Bolero-Songs Ruined Me) , Pregones Theatre, Bronx, NY. See also:

[www.pregones.org/Eng/productions/bolero.html](http://www.pregones.org/Eng/productions/bolero.html)] Here's a more traditional, really wonderful performance.

About 60 have been converted so far, and we're gaining speed.

[“Lessons along the way” Slide] Lessons learned: This Mac G4 screen [circle with slash through it] means “I’m dead and I’m never coming back”, which happened in May, 2004. A G5 replaced it. We demand intense processing, so it’s no surprise to go through equipment. The Medéa RAID should have worked well with the G4, but didn’t. Companies that seemed viable went down, particularly on the derivatives side. We expect more such architectural issues when we upgrade to HD.

[another slide, bullets continued] We’re preserving SD signal, storing to tape for now. We got our workflow worked out end-to-end. We’ve got a non-homogenous collection. We can ingest a number of tape formats - we concentrate on Mini-DV and VHS, but now have a U-matic deck (thanks to Preservation). We’ve gained a great deal of in-house knowledge about equipment and metadata, so we can do a better job for NYU.

[“Resources” slide] For metadata: LoC, SMPTE. For preservation: AMIA [Others on slide: SMIL, Doom9, Ars Technica, NYU-TV.]

#### **Q & A [“with tough question to Rick”]**

*Q.* You’ve concentrated on preservation of the signal. Does this give you an option to transform this signal to other formats, e.g., lossless?

*A.* We are looking at other formats. Like a lot of archivists, we at NYU are looking at open source and what’s doable now. Some of JPEG2000 family looks good for the future. If we can wrap it up in another wrapper, that’s fine, as long as the transform is reversible.

*Q.* Since Sony’s DigiBeta is a proprietary format, when you want to convert, what information are you retaining?

*A.* It’s the “largest” tape container, and output is SDI, which is a standard. DigiBeta is widely used by production houses, so won’t go away. (There was a back-and-forth about the analogous TIFF situation, which is also proprietary. It’s “open” but could be policed by Adobe at any time.)

*Q.* Why up-convert DV to DigiBeta, as opposed to staying with native DV, e.g., DVCam? The file just gets bigger going to DigiBeta.

*A.* Because we wanted uniformity of the master file, and expect eventually to move it to disk. That is easier to manage than five different formats of both media and metadata.

*Questioner responds.* In a similar situation, we chose three master formats - DV, DigiBeta, and HDCam.

*C.* Returning to the issue of the proprietary nature of DigiBeta.... As someone who has worked on the relevant SMPTE committees, I’m concerned about preserving both content and the mechanism to read that content. With DigiBeta, it’s the machine that’s of

concern. With disk-based compression, it's the algorithm. With compression systems, do choices made today impact our ability in 100-150 years to get the data back?

A. We're just trying to get to 2010 with DigiBeta, then store info on cheap disk.

*Commenter responds.* Will you compress the SDI on disk, or leave it uncompressed?

A. Leave as uncompressed as possible. If we compress, we'll have a cabinet to store the codec spec for playback, so an engineer 100 years from now can re-create the mechanisms.